

CLIP BASED TRICK MODES

The invention herein is related to the field of trick play modes for presentation of
5 stored multimedia performances. These multimedia performances usually consist of
audio, video and optionally shown text captions. A performance usually has a normal
forward speed at which the performance is usually presented to the audience. Recording
and playing devices for stored performances often provide, trick play modes for video
such as fast foreword, reverse and fast reverse play so that the user can skip material or
10 more conveniently find a desired position to start recording or playing a performance.
Trick play video modes are usually at different speeds or in reverse direction and often at
different smoothness then normal play. During known trick video play audio and captions
is not played.

During normal play, video frames are typically presented to the audience at a
15 constant predetermined video frame rate with each frame presented for a constant
predetermined normal frame period. Similarly, digital audio is usually divided into audio
frames which are also presented to the audience at a constant predetermined audio frame
rate. This results in the presentation having a normal speed for video and audio. On the
other hand, data for captions is presented at a varying rate that is much lower than the
20 video or audio frame rate. During normal play each caption is presented for a normal time
period that is long enough for the audience to read the caption. Usually during normal
play, each caption stays on the screen until it is erased by the next caption or blank
caption.

The speed and direction of a trick play mode is usually expressed in relation to the
25 normal forward speed at which the performance is normally presented to an audience. For

example, 2X speed means that the video is being presented at twice the normal speed. To provide 2X video speed the video frames may be presented at twice the normal video frame rate for half the normal video frame period. However, video displays often only accept frames at a single predetermined frame rate. In that case 2X speed may be

5 achieved by skipping half the frames and presenting the remaining frames at the normal frame rate and for the normal frame period. Presenting half the frames at twice the frame rate may result in a visibly jerky motion. As trick speed video increases beyond about 8X then the normal movement of objects and people and panning of the camera and the changing of camera angles will result in a presentation can not be understood by the

10 audience.

MPEG is a common format for encoding to compress digital video. For example, there may be, for example, 16 pictures in each GOP. Each picture is called a frame. The first frame presented in a GOP is called an I-frame and its decoding is independent (like a JPEG picture) it does not depend on the decoding of any other frame. Other frames in a

15 GOP are P-frames and B-frames. The decoding of the P-frames depend on the decoding of the closest previous I-frame or P-frame in the GOP. The decoding of the B-frames depends on the decoding of both the closest subsequent P-frame as well as the decoding of the closest previous I-frame or P-frame.

The MPEG data stream consists of a series of packets each having a header with a predetermined format and each having a data section. The header includes an ID which

20 identifies the packets containing the data for a particular video frame. All the data packets for one video frame is transmitted/stored before any of the data packets for the following video frame so that the packets for different video frames are not intermixed. However,

the data packets for other media such as audio frames or captions may be intermixed with the data packets for the video frames.

The compressed data for the video frames is transmitted and stored in the order in which it is needed for decoding so that a GOP that would be presented in the order of:

5 I,B₁,B₂,P₁,B₃,B₄,P₂,B₅,B₆,P₃,B₇,B₈,P₄,B₉,B₁₀,P₅

would be transmitted and stored in the order of:

I,P₁,B₁,B₂,P₂,B₃,B₄,P₃,B₅,B₆,P₄,B₇,B₈,P₅,B₉,B₁₀

because, for example, the data of the P₁ frame is required for the decoding of the B₁ and B₂ frames that are presented before the P₁ frame is presented.

10 Providing high speed trick play modes for MPEG compressed video has been challenging because decoding requires substantial processing effort. For example, skipping half the frames does not proportionately decrease the required decoding effort since all the I-frames and P-frames still have to be decoded in order to decode the required B-frames, and the I-frames and P-frames require much more effort to decode
15 than the decoding of the B-frames.

In a well known fast foreword trick play mode for MPEG compressed digital multimedia presentations, only the I-frames and none of the other types of frames are presented. If the multiple of the speed of the trick mode presentation is less than the number of I-frames in a GOP, then I-frames have to be presented for a longer than normal
20 frame period. This results in an extremely jerky presentation, but at least the audience can recognize portions of the video at in the fast foreword trick play mode. For example, for 4X speed and 16 pictures in each GOP, each I-frames would be shown for 4X the normal frame time. The beginnings of the I-frames are usually called video trick mode entry points because decoding can only begin at an I-frame. In order to quickly find the I-

frames an indication of the locations of the video trick mode entry points (e.g. the beginnings of the I-frames) can be provided, such as, by providing a flag in the last data packet before an I-frame. A video trick mode exit point is a position in the MPEG data stream where decoding can stop without leaving any errors or anomalies called artifacts.

- 5 The end of each video frame is a trick mode exit point for MPEG video.

In order to minimize confusion, applicant has used the term "media" to refer to different types of presentations such as audio, video, text. Applicant has used the term "medium" to refer to storage medium such as record carrier and to transmission medium such as space or wires or optical cables through which signals in the form of
 10 electromagnetic waves are transmitted. Those who are unfamiliar with the art should be careful since both media and medium are often used interchangeably in the field.

In US5664044 to Ware, both audio and video are presented in a faster than normal trick play mode. In US6064794 to McLaren et. al., determining entry points for trick play modes of MPEG encoded video is disclosed. In WO99/65239 to Erenburg et. al. storing
 15 entry points for trick mode play along with MPEG encoded video is disclosed. Those skilled in the art are directed to the above references which are hereby incorporated in whole by reference.

In the invention herein, a clip based trick mode for playing a performance is provided. The performance may be, for example, an audio and/or video performance. A
 20 performance includes a multitude of sequential frames for presentation during normal playing of the performance in a predetermined normal frame presentation order at a predetermined normal frame rate. The performance also has a normal speed of motion and a normal direction at which the presentation is normally presented to the audience. The performance is stored on a storage medium and then portions of the performance are

read from the storage medium for playing the performance. Frames of the performance are subdivided into trick play clips and fast skim clips. The trick play clips and the fast skim clips each contain multiple sequential frames with the trick play clips alternate with the fast skim clips in the normal frame presentation order. The frames of different media
 5 may be included in the same trick play clips or the frames of different media may be subdivided differently.

During the playing of the performance in the trick play mode, the trick play clips are played and the fast skim clips are skimmed. The skimming may be either skipping of fast skim clips entirely or playing fast skim clips at a substantially higher speed than the
 10 trick play clips. Some frames of the trick play clips may need to be skipped for high speed trick play video modes. If the fast skim clips are played at a high speed then some of the frames of the fast skim clips may have to be skipped. The average speed of the playing and skimming in the trick play mode is substantially different than the normal speed for the performance. The trick play clips are sufficiently long and presented at a sufficiently
 15 low speed so that the content of the trick play clips can be understood by a human audience.

In the trick play mode, the speed at which the trick play clips are played may be normal or may be slower or faster than normal and may be either forward or reverse. The speed for playing the trick play clips may be predetermined during editing or user
 20 adjustable during the trick mode play. For a multimedia performance, in the invention, the speed at which different media are played, may be different for different media. The trick play clips for video may be different than the trick play clips for audio. For example, a trick play mode could be provided with 2X audio speed and 8X video speed in which, all the clips will be used for video, but only 1/4 of the clips will be used for audio. This

allows a fast smooth video playback of selected clips and simultaneous audio playback of selected clips that are still understandable. The order in which the clips are presented can be different than the order in which the frames within each clip are shown. One possibility is a fast reverse video trick mode in which clips are played in reverse order
 5 with the video frames of the clips played in reverse order, but the audio frames of the clips played in forward order so the audio can be understood. The length of the trick play clips and the length of the fast skim clips may be the same length or different lengths and the lengths may be predetermined or user adjustable during the trick mode play.

The trick play clips may begin at selected trick play entry points. When trick play
 10 clips are preselected during the editing process, then in addition to indicating the trick play entry points, an indication of which trick play entry points are selected for the start of trick play clips are also stored. The indication of the starting positions of the trick play clips may be indicated within the performance or in a table outside of the performance.

The above and other objects and features of the invention will become apparent to
 15 those skilled in the art from the following detailed description presented with reference to the following drawings:

Figure 1 is an example of a flow diagram of the method of the invention.

Figure 2 illustrates an example signal of the invention.

Figure 3 shows additional details of the example signal of figure 2.

20 Figure 4 illustrates a recorder of the invention.

Figure 5 illustrates a player of the invention.

Figure 6 illustrates a digital video cassette recorder of the invention.

Figure 1 illustrates the method of the invention. In step 100 a performance is provided. A performance includes a multitude of sequential frames for presentation

during normal playing of the performance in a predetermined normal frame presentation order at a predetermined normal frame rate. The presentation has a normal speed and a normal direction of play for presentation to an audience. For example, the performance may be an audio performance or a video performance or a multimedia performance with audio, video and text captions. The performance may be broadcast or it may be received interactively.

In step 102, trick play clips for a trick mode are selected. The trick play clips contain multiple frames and the trick play clips are mutually separated by multiple frames that are defined as fast skim clips. Clips for multiple different trick modes may be selected during a authoring process of the performance (as shown) or alternatively, the trick play clips can be selected automatically, as needed, during the playing in a trick mode. Automatic selection is discussed in more detail below in relation to step 112. If the trick play clips are selected during the authoring process, they can be selected so that, the trick play clips correspond to portions that are logically understandable and recognizable. For example, trick play audio clips can be selected that contain an understandable phrase or statement, trick play video clips can be selected that contain recognizable portions of the performance. The trick play clips may even be highlights or important or memorable portions of scenes.

For different trick play modes, different trick play clips may be selected or the same set of trick play clips can be used for different trick play modes. For different media, such as video and audio, different trick play clips can be selected or the same trick play clips can be used for different media. For one media such as audio, the trick play clips can be preselected during the editing process, and for another media, such as video, the trick play clips, can be automatically selected during play in the trick play mode,

based on, for example, trick mode entry points and user selected clip lengths. The trick play clips may all be the same uniform size or they may be different sizes so that the clips are more understandable.

In a video presentation using MPEG compression, usually only the beginning of each I-frame (the beginning of the first packet of the I-frame) is a video trick mode entry point. That is, video decoding can only begin at the beginning of each I-frame. Thus, trick mode clips for video must start prior to or at the beginning of an I-frame. Generally, it is convenient if the trick play clip that is used for video starts at a video trick mode entry point and ends at a trick mode video exit point. The end of each video frame (the end of the last packet of the frame) is a video trick mode exit point.

In MPEG audio, each audio frame is independently decodable so that the beginning of each audio frame (the beginning of the first packet for the audio frame) is an audio trick mode entry point and the end of each audio frame (the end of the last packet for the audio frame) is an audio trick mode exit point.

In step 104, the performance is compressed. For example, MPEG II type video compression may be used for video and MPEG type audio compression may be used for audio. Alternatively, other known methods of compressing of performances could be used. The compression may be variable, such as MPEG II, or the compression can be a fixed ratio type compression. In MPEG compressed performances, the beginning of each I frame is a video trick mode entry point and the beginning of each audio frame is an audio trick mode entry point, and the beginning of each caption is a text trick mode entry point.

In step 106, indication of the position of the trick play clips are provided. The indication may be provided as pointers to the positions of the start of the trick play clips,

or the positions can be indicated within the compressed performance. For example, a tag in the header of the last packet of an MPEG II compressed stream prior to the start of the trick play clip can be used to indicate the beginning of a trick play clip.

For variable compression schemes such as MPEG II, generally indications of
5 positions can not be determined until after the performance is compressed. For MPEG video compression, the beginning of each group of pictures is a video trick mode entry point, and for convenience in decoding, the start of each trick play video clip may correspond to one of these video trick mode entry points.

It may be convenient to provide indications of both the video trick mode entry
10 points and the positions of the video trick play clips or indications of which video trick mode entry points are also the beginnings of video trick play clips.

In step 108, the performance is stored on a storage medium. If the positions of the trick play clips are predetermined, then indications of the positions of the trick play clips are also stored on the same storage medium. The position indication for the trick play
15 clips may be in the form of pointers in which case they can be stored on the storage medium as a table of pointers. If the positions indication is in the form of tags in the packets of an MPEG II compressed performance then the packets are modified prior to storing on the storage medium.

If trick mode entry points are indicated in the stored performance and if the trick
20 play clips begin at selected trick mode entry points, then an indication of which trick mode entry points are also the beginnings of trick play clips is also stored so as to indicate the beginnings of the trick play clips.

In step 110, in a trick play mode, the trick play clips are read from the storage medium. If the positions of the trick play clips are predetermined, then indications of the

positions of the trick play clips are also read from the storage medium. If the fast skim clips are going to be skipped in the trick play mode then at least portions of the fast skim clip may not have to be read from the storage medium. If the indications are contained in the compressed performance then portions of the fast skim clips has to be read to find and

5 identify the indications. For example, if there are tags in the headers of packets before or at the start of the trick play clips, then at least a portion of a fast skim clip just before the start of the trick play clip has to be read in order to read the indications of the start of the trick play clip. Otherwise, if pointers to the trick play clips are stored on the storage medium, (e.g. in a table) then the fast skim clips do not need to be read, and the trick play

10 clips can be accessed directly.

In step 112 during the playing of the performance in the trick play mode, the trick play clips are played and the fast skim clips are skimmed. The skimming may be either skipping of the fast skim clips entirely or playing the fast skim clips at a substantially higher speed than the trick play clips.

15 In the trick play mode, the speed at which the trick play clips are played may be normal or may be slower or faster than normal speed and may be either forward or reverse. The speed for playing the trick play clips may be predetermined during editing or user adjustable during the trick mode play. For a multimedia performance, the speed may be different for different media. The trick play clips for video may be different than the

20 trick play clips for audio.

The average speed of the playing and skimming in the trick play mode is substantially different than the normal speed. The trick play clips are sufficiently long and presented at a sufficiently low speed so that the content of the trick play clips can be

understood by a human audience. The fast skim clips may be played at such a high speed that much of the fast skim clip can not be understood or even recognized by the audience.

In an example trick play mode, the trick play video clips contain 8 frames corresponding to half of a group of pictures in an MPEG compression scheme. The fast skip clips may contain 24 frames corresponding to 1-1/2 groups of pictures. The trick play video clips may be played at 4X speed by displaying each of the video frames of the trick play clip for 1/4 of the normal frame period, and the fast skim clips may be played at a much higher speed (e.g. 16X speed) by only playing each I-frame in the fast skim clip for the normal frame period.

The direction of the clip-to-clip order in the trick mode can be different than the direction of the order in which the frames of each clip are played. For example, if clips 1, 2, 3 in normal order each have frames A, B, C, in normal order then the clips can be played in reverse clip order of 3, 2, 1 while the frames within each clip is played in the normal order of A, B, C. Thus the clips are played in reverse clip order while the frames of the clips are played in normal order. This allows audio to be understood while the presentation is being played in a reverse trick mode.

The length of the trick play clips and the length of the fast skim clips may be the same length in relation to normal playing time or different lengths and the lengths of the clips may be predetermined or user adjustable during the trick mode play.

Figure 2 illustrates a specific example embodiment of the signal of the invention. A first portion 122 may be provided which contains indications of the positions of the trick play clips. The remaining portions of the signal may contain multiple trick play clips 123, 125 and 127 alternating with multiple fast skim clips 124, 126, and 128. The ellipse following fast skip clip 128 indicates that the alternating clips continue. For example, a

one hour performance with 1/4 second long clips may contain about 7,000 trick play clips.

The first clip 123 following the table of pointers is a trick play clip, but alternatively (not shown) the performance may begin with a fast skip clip. The indications
 5 of position of portion 122 may be a table containing pointers to the positions of the start of each trick play clip or a list of pointers. Pointers to the ends of the trick play clips (beginnings of the fast skim clips) may also be required if the trick play clips are not uniform in size. Instead of providing the pointers in a separate portion before or after the performance, pointers for subsequent trick play clips may be provided within previous
 10 trick play clips. Alternatively, indications such as a clip ID can be provided in the compressed performance at the beginning of a trick play clip or near the end of the previous fast skip clip, and the last portion of the fast skip clip can be scanned to find the beginning of the trick play clip.

For a video performance (i.e. a performance that includes video) each trick play
 15 clip will include data for a multitude of video frames. The number of frames contained in a trick play clip will be sufficient so that when the clip is played in the trick play video mode, the audience will be able to perceive some of the contents of the clip. That is, when played in the trick play mode, the trick play clips will be more just be an unintelligible blur, but would be recognizable by someone who was very familiar with the performance.

20 An advantage of providing pointers in the signal prior to the trick play clips is that when the signal is stored and the playing of the fast skim clips are going to be skipped entirely, then the reading and processing of the fast skim clips can also be skipped.

Figure 3 is a more detailed illustration of portions of the example signal of figure 2. Each clip 123, 124 and 125 contains a multitude of respective packets such as packets 132, 135 and 138 respectively and each packet contains a respective header 133, 136 and 139 and each packet includes a respective data section 134, 137 and 140. In addition or in the alternative to the other indications, the locations of the trick play clips may be indicated in the last packet of the previous fast skim clip, for example, a flag or clip ID may be provided in header 136 to indicate that packet 138 is the first packet of a fast play clip. Also, in addition or in the alternative to other indications, the locations of the trick play clips may be indicated in the first packet of the trick play clips such as by a flag in header 139 of the first packet 138 of trick play clip 125. Also, in addition or in the alternative to other indications, the locations of the trick play clips may be indicated in the previous trick play clip, such as, by a pointer in a packet of trick play clip 123 to the start 129 of the following trick play clip 125.

In this example, the trick play clips are the same for every type of media that is played during the trick play mode. However, playing of each type of media can only begin at a trick mode entry point for that type of media. It is possible to arrange the packets of the performance so that, at the start of each trick play clip, the next packet for each type of media to be played is a trick mode entry point for that type of media. Alternately, after the start of the trick play clip, playing of each type of media will just be delayed until a trick mode entry point for that type of media is reached. Similarly, it is possible to arrange the packets so that the last packet of each type of media in the trick play clip is a trick mode exit point so that the trick mode can stop playing without causing artifacts. Alternatively, play of each media can end at the last trick mode exit point for that type of media in the trick play clip. Another alternative is for trick mode playing of

each type of media (being played), to end when the first trick mode exit point is reached after the end of the trick mode clip.

Alternately, different trick play clips can be defined for each different type of media. In this case indications will be provided for the positions of the trick mode clips of each type of media. If the playing of the fast skim clips will be skipped in the trick mode, then it is possible to reduce the reading of fast skim clips by selecting the trick play clips for different media so that at least in some cases the different fast skim clips overlap for all the types of media.

Figure 4 illustrates a recorder 200 of the invention for producing storage medium containing the performance including the trick play clips and indications of the locations of the trick play clips. A performance is provided at input 202. Selection unit 204 is used to select the trick play clips of the performance for use in trick play modes. Encoder 206 compress the performance. Indication determining unit 208 determines indications of the positions of the trick play clips to be stored with the performance. Formatting unit 210 formats the performance and the indications of the positions of the trick play clips together into a signal 211 of the invention which is transmitted through a transmission medium to write head 212 which records the signal onto storage medium 214.

Encoder 206 is not required if the performance does not have to be encoded. If storage medium 214 is a solid state medium then write head 212 is not required and the signal can be directly stored on the solid state medium. Storage medium 214 may for example be a removable record carrier such as a DVD disc or a master disc used in a press for subsequently producing multiple record carriers.

The

Figure 5 illustrates a player 240 of the invention for playing the performance including playing the trick play clips in the trick play mode. Storage medium 214 is read by read head 242 which is controlled by controller 244. If the trick play clips are predetermined and if the fast skim clips are skipped during the trick play mode, then

5 controller 244 receives the indications of the positions of the trick play clips and during the trick play mode, the controller controls the read head to skip reading of the fast skim clips. Selector 246 selects portions of the read signal to be decoded depending on a signal from controller 244. If the trick play clips are not predetermined then during the trick play mode the selector is used to select the trick play clips and fast skip clips for playing in the

10 trick play mode. Decoder 248 decompresses the selected portions of the read signal and provided the decompressed signal 249 to frame formatter 250 which formats the signal into presentation signal 251 which is transmitted to presentation unit 252 which plays the presentation to an audience in the normal and trick play mode. The formatting of the frame formatter 250 may be under the control of controller 244 to provide only the

15 desired frames and control presentation unit 252 to provide the desired frame rate.

If the trick play clips are predetermined and if the fast skim clips are skipped during the trick play mode, then selector 426 may not be required.

The recorder of figure 4 can be combined with the recorder of figure 5 to produce a single apparatuses having both recording and playing capabilities.

20 figure 6 illustrates a digital video cassette recorder 270. The recorder has an opening 271 for receiving a video cassette 272 containing a video tape 273. A drive 274 rotates a rotating head 275 reads frames of a digital video performance from slanted tracks across the video tape. The video performance is encoded based on groups of pictures (GOPs) and the beginning of a GOP is at the start of each track. A tape drive

mechanism 276 moves the video tape longitudinally at the rotating head. A controller 277 contains programmed modules for: controlling the speed of rotation of the head and the linear speed of the tape in synchronization to provide a normal speed of reading consecutive frames from one of the tracks and a different trick play mode speed for

5 reading consecutive frames from one of the tracks at a different speed; and controlling the playing of portions of the performance in a trick play mode in which multiple frames in a first portion of the track beginning at the start of the track are played and multiple frames of a second portion of the track ending at the end of the track are skimmed, the second portion being as long or longer than the first portion, the skimming being skipping or

10 playing at a faster rate than the first portion is played, and the speed at which frames are played is different than the speed at which frames are read from the tape.

The invention has been disclosed with reference to specific preferred embodiments, to enable those skilled in the art to make and use the invention, and to describe the best mode contemplated for carrying out the invention. Those skilled in the

15 art may modify or add to these embodiments or provide other embodiments without departing from the spirit of the invention. Thus, the scope of the invention is only limited by the following claims: